# Standard D eviation and N orm al D istribution

For questions in the Q uantitative C om parison form at (“Q uantity A ” and “Q uantity B ” given),the answ er choices are alw ays as follow s:

(A ) Q uantity A is greater.
(B ) Q uantity B is greater.
(C ) The tw o quantities are equal.
(D ) The relationship cannot be determ ined from the inform ation given.

For questions follow ed by a num eric entry box ☐,you are to enter your ow n answ er in the box.For questions follow ed by fraction-style num eric entry boxes $\frac{☐}{☐}$,you are to enter your answ er in the form of a fraction.Y ou are not required to reduce fractions.For exam ple,if the answ er is 1/4,you m ay enter 25/100 or any equivalent fraction.

A ll num bers used are real num bers.A ll figures are assum ed to lie in a plane unless otherw ise indicated.G eom etric figures are not necessarily draw n to scale.Y ou should assum e,how ever,that lines that appear to be straight are actually straight,points on a line are in the order show n,and all geom etric objects are in the relative positions show n.C oordinate system s,such as *xy*-planes and num ber lines,as w ell as graphical data presentations such as bar charts,circle graphs,and line graphs, *are* draw n to scale.A sym bol that appears m ore than once in a question has the sam e m eaning throughout the question.

1.Set *S* {5,10,15}

If the num ber 15 w ere rem oved from Set *S* and replaced w ith the num ber 1,000,w hich of the follow ing w ould change?

Indicate <u>all</u> such statem ents.

☐ The m ean
☐ The m edian
☐ The standard deviation

2.

Set *W* : -9,-3,3,9
Set *X* : 2,4,6,8
Set *Y*: 100,101,102,103
Set *Z*: 7,7,7,7

W hich of the follow ing choices lists the four sets above in order from sm allest standard deviation to greatest standard deviation?

(A ) *W* ,*X* ,*Y*,*Z*
(B ) *W* ,*Y*,*X* ,*Z*
(C ) *W* ,*X* ,*Z*,*Y*
(D ) *Z*,*Y*,*X* ,*W*
(E) *Z*,*X* ,*Y*,*W*

3.

Set *N* is a set of *x* distinct positive integers w here $x > 2$.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The standard deviation of Set *N* | The standard deviation of Set *N* if every num ber in the set is m ultiplied by -3 |

4.Set *S* is a set of distinct positive integers.The standard deviation of Set *S* m ust increase if w hich of the follow ing w ere to occur?

Indicate all such statem ents.

☐ Each num ber in the set is m ultiplied by 1/2.
☐ The sm allest num ber is increased to becom e equal to the m edian.
☐ The sm allest num ber is increased to becom e larger than the current largest num ber.
☐ The largest num ber is doubled.

5.The 75th percentile on a test corresponded to a score of 700,w hile the 25th percentile corresponded to a score of 450.

| **Q uantity A** | **Q uantity B** |
|---|---|
| 800 | A 95th percentile score |

6.Set *X* consists of 9 total term s,but only tw o different term s.Six of the term s are each equal to tw ice the value of each of the rem aining 3.W hich of the follow ing w ould provide sufficient additional inform ation to determ ine the average of the set?

Indicate all such statem ents.

☐ The sm aller num ber is positive and is 3 less than the larger num ber.
☐ The standard deviation of the set is equal to $2\sqrt{3}$.
☐ The biggest term in the set is 6.

7.Set *S* = {2,5,7,11,16,24,28,50,52,101,120,130}

W hat is the average of the first quartile (“Q 1”) and the third quartile (“Q 3”) of set *S*?

(A ) 9

(B ) 26
(C ) 42.75
(D ) 76.5
(E) 85.5

8.The test scores at M illbrook H igh School are norm ally distributed,and the 60th percentile is equal to a score of 70.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The 30th percentile score | 35 |

9.The lengths of a certain population of earthw orm s are norm ally distributed w ith a m ean length of 30 centim eters and a standard deviation of 3 centim eters.O ne of the w orm s is picked at random .

| **Q uantity A** | **Q uantity B** |
|---|---|
| The probability that the w orm is betw een 24 and 30 centim eters,inclusive | The probability that the w orm is betw een 27 and 33 centim eters,inclusive |

10.The hourly w age paid to w orking adults in M aplew ood is norm ally distributed around a m ean of $18 per hour w ith a standard deviation of $3.50.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The percent of w orking adults in M aplew ood w ho are paid betw een $18 and $25 per hour,inclusive | 40% |

11.H om e values am ong the 8,000 hom eow ners of Tow n X are norm ally distributed,w ith a standard deviation of $11,000 and a m ean of $90,000.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The num ber of hom eow ners in Tow n X w hose hom e value is above $112,000 | 300 |

12.Exam grades am ong the students in M s.H arshm an's class are norm ally distributed,and the 50th percentile is equal to a score of 77.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The num ber of students w ho scored less than 80 on the exam | The num ber of students w ho scored greater than 74 on the exam |

13.The length of bolts m ade in factory *Z* is norm ally distributed,w ith a m ean length of 0.1630 m eters and a standard deviation of 0.0084 m eters.The probability that a random ly selected bolt is betw een 0.1546 m eters and 0.1756 m eters long is betw een

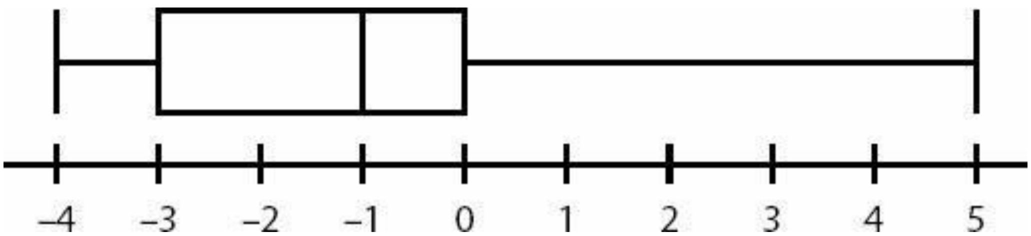
(A ) 54% and 61%
(B ) 61% and 68%
(C ) 68% and 75%
(D ) 75% and 82%
(E) 82% and 89%

14.B irth w eight of babies born at C ity H ospital is norm ally distributed.A baby 2 standard deviations above the m ean birth w eight w eighs 10.8 pounds,and a baby 1 standard deviation below the m ean birth w eight w eighs 5.85

pounds.

| Quantity A | Quantity B |
|---|---|
| Twice the weight of a baby 2 standard deviations below the mean | The weight of a baby 1 standard deviation above the mean |

15.Which of the following sets of data applies to this graph?




(A) -4,-4,-2,0,0,5
(B) -4,1,1,3,4,4
(C) -4,-4,-3,1,5
(D) -5,3,4,5
(E) -4,-4,-2,-2,0,0,0,5

16.If a set of data consists of only the first ten positive multiples of 5,what is the interquartile range of the set?

(A) 15
(B) 25
(C) 27.5
(D) 40
(E) 45

17.On a given math test out of 100 points,the vast majority of the 149 students in a class scored either a perfect score or a zero,with only one student scoring within 5 points of the mean.Which of the following logically follows about Set *T*,made up of the scores on the test?

Indicate all such statements.

☐ Set *T* will not be normally distributed.
☐ The range of Set *T* would be significantly smaller if the scores had been more evenly distributed. ☐ The mean of Set *T* will not equal the median.

18.If Set *X* is a normally distributed set of numbers with a mean of 4 and a standard deviation of 4,approximately what is the probability that a number chosen at random from the set will be negative?

(A) 1/10
(B) 1/6
(C) 1/4
(D) 1/3
(E) 1/2

19.Jane scored in the 68th percentile on a test,and John scored in the 32nd percentile.

| Quantity A | Quantity B |
|---|---|

The proportion of the class that received a score less than John’s score

The proportion of the class that scored as high as or higher than Jane

20.If a set of data has a m ean of 4.2 and a standard deviation of 7.1,w hat is the range of values that lie w ithin 2 standard deviations of the m ean?

(A ) -2.9 to 11.3
(B ) -2.9 to 12.6
(C ) -10 to 12.6
(D ) -10 to 18.4
(E) 4.2 to 18.4

21.If octiles divide up a set of data into 8 ordered groups,each w ith the sam e num ber of term s,w hat is the m edian of the sixth octile group of the set of data com posed of the integers from 25 to 48,inclusive?

22.In a class w ith 20 students,a test w as adm inistered,scored only in w hole num bers from 0 to 10.A t least one student got every possible score,and the average w as 7.

**Q uantity A**

4

**Q uantity B**

The low est score that tw o students could have received

23.

| | | | |
|---|---|---|---|
| Frequency | 6 | 5 | 5 |
| Result | 4 | 6 | 8 |

**Q uantity A**

The m ode of this data set

**Q uantity B**

5

24.A test is scored out of 100 and the scores are divided into five quintile groups.Students are not told their scores, but only their quintile group.

**Q uantity A**

The scores of tw o students in the bottom quintile group,chosen at random and added together

**Q uantity B**

The score of a student in the top quintile group,chosen at random

25.In a set of 10 m illion num bers,one percentile w ould represent w hat percent of the total num ber of term s?

(A ) 1,000,000
(B ) 100,000
(C ) 10,000
(D ) 100

(E) 1

26.W hat is the range of the set of num bers com prised entirely of {1,6,$x$,17,20,$y$} if all term s in the set are positive integers and $xy$ = 18?

(A ) 16
(B ) 17
(C ) 18
(D ) 19
(E) C annot be determ ined from the inform ation given.

27.O n a particular test w hose scores are distributed norm ally,the 2nd percentile is 1720,w hile the 84th percentile is 1990.W hat score,rounded to the nearest 10,m ost closely corresponds to the 16th percentile?

(A ) 1,750
(B ) 1,770
(C ) 1,790
(D ) 1,810
(E) 1,830

28.A data set contains at least tw o different integers.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The range of the data set | The interquartile range of the data set |

29.In a norm ally distributed set of data,one standard deviation above the m ean is 77 and the standard deviation is 10. W hat is the m ean of the data?

30.Som e rock sam ples are w eighed,and their w eights are determ ined to be norm ally distributed.O ne standard deviation below the m ean is 250 gram s and one standard deviation above the m ean is 420 gram s.

| **Q uantity A** | **Q uantity B** |
|---|---|
| 335 gram s | The m edian w eight,in gram s |

31.In a norm ally distributed set of data,the m ean is 12 and the standard deviation is less than 3.

| **Q uantity A** | **Q uantity B** |
|---|---|
| N um ber of data points in the set located betw een 9 and 15 | 60% of the total num ber of data points |

32.

| **Q uantity A** | **Q uantity B** |
|---|---|
| The standard deviation of the set 10,20,20,20,20, 20,30 | 10, The standard deviation of the set 20,30 |

33.

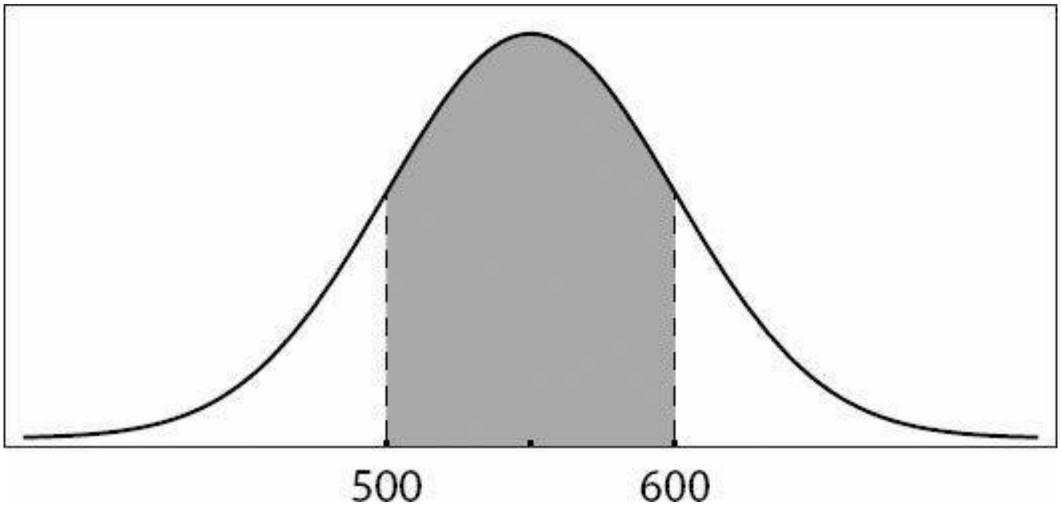
**Quantity A**

The standard deviation of a set of numbers with a range of 8

**Quantity B**

The standard deviation of a set of numbers consisting of 3 consecutive multiples of 3

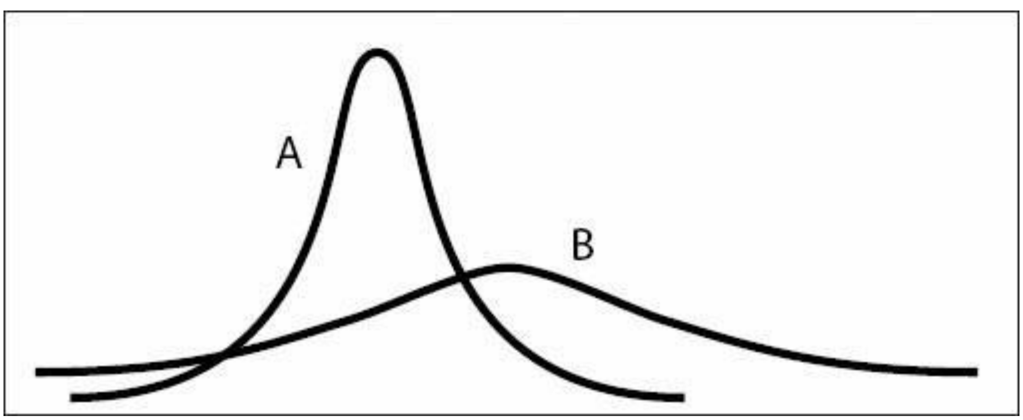
34.




The graph represents the normally distributed scores on a test.The shaded area represents approximately 68% of the scores.

| **Quantity A** | **Quantity B** |
| --- | --- |
| The mean | 550 |

35.




A and B are graphical representations of normally distributed random variables *X* and *Y*,respectively,with relative positions,shapes,and sizes as shown.Which of the following must be true?

Indicate all such statements.

- ☐ *Y* has a larger standard deviation than *X* .
- ☐ The probability that *Y* falls within 2 standard deviations of its mean is larger than the probability that *X* falls within 2 standard deviations of its mean.

☐ *Y* has a larger mean than *X* .

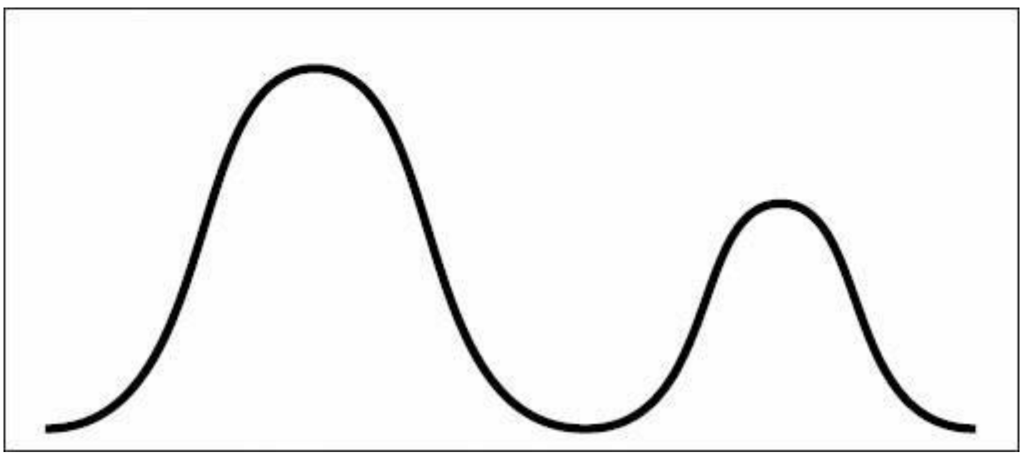
36.300 test results are integers ranging from 15 to 75,inclusive.Dominick's result is clearly in the 80th percentile of those results,not the 79th or the 81st.

| **Quantity A** | **Quantity B** |
| --- | --- |
| Number of other test results in the same percentile as Dominick's | Maximum number of other test-takers with the same result as Dominick |

37.The outcome of a standardized test is an integer between 151 and 200,inclusive.The percentiles of 400 test scores are calculated,and the scores are divided into corresponding percentile groups.

| **Quantity A** | **Quantity B** |
| --- | --- |
| Minimum number of integers between 151 and 200, inclusive,that include more than one percentile group | Minimum number of percentile groups that correspond to a score of 200 |

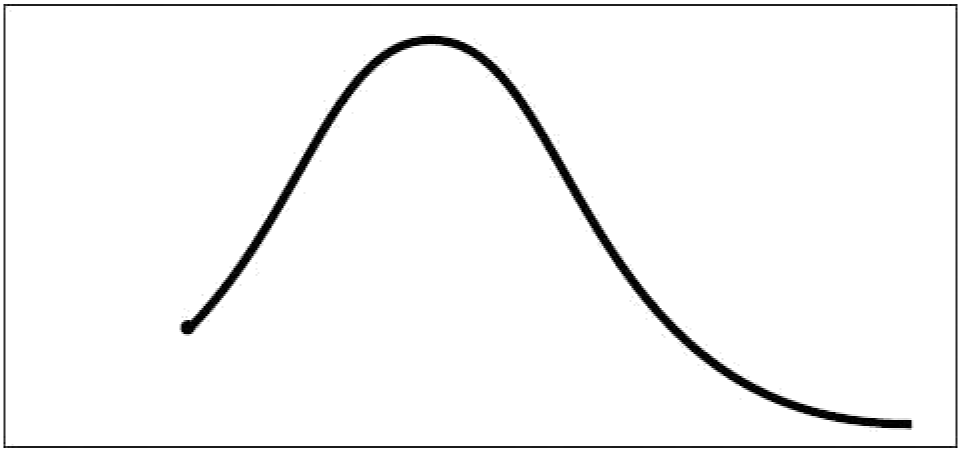
38.



Which of the following would the data pattern shown best describe?

(A ) The number of grams of sugar in a selection of drinks is normally distributed.
(B ) A number of male high school principals and a larger number of female high school principals have normally distributed salaries,distributed around the same mean.
(C ) A number of students have normally distributed heights,and a smaller number of taller,adult teachers also have normally distributed heights.
(D ) The salary distribution for biologists skews to the left of the median.
(E) The maximum-weight bench presses for a number of male athletes are normally distributed,and the maximum-weight bench presses for a smaller number of female athletes are also normally distributed,although around a smaller mean.

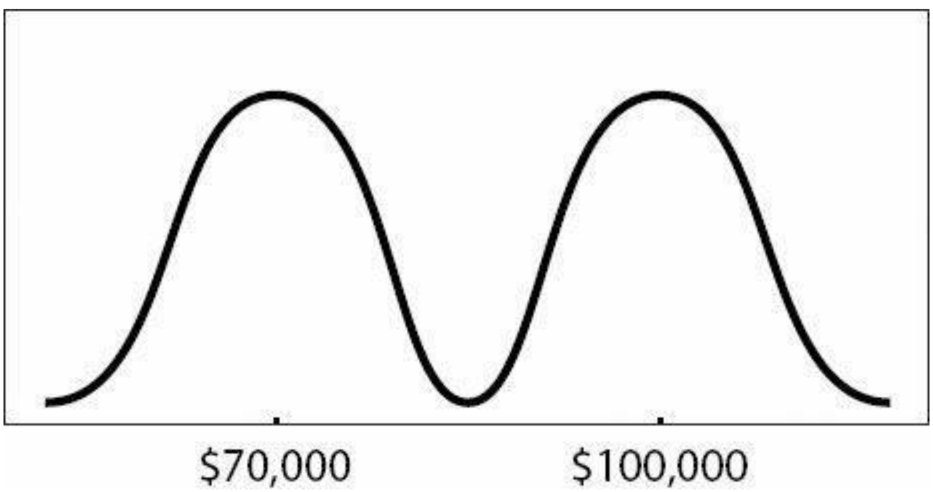
39.



W hich of the follow ing w ould the data pattern show n best describe?

(A ) The w eights of raccoons in a population are norm ally distributed.
(B ) Salaries in a certain field appear norm ally distributed,except that salaries cannot dip below the lim its of a m inim um -w age law .
(C ) The fraction of people at a certain age in a certain population is inversely proportional to age. (D ) a set of consecutive integers
(E) a set w ith a standard deviation of zero

40.

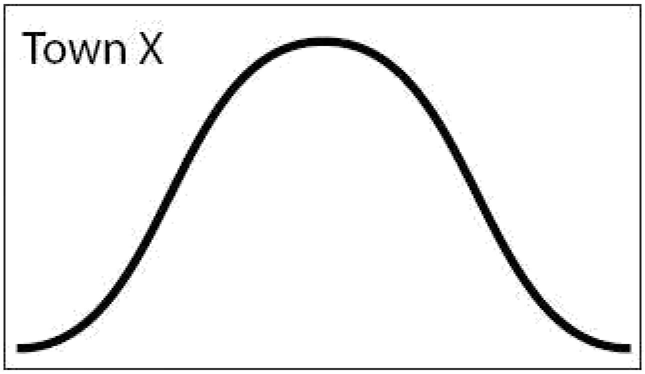


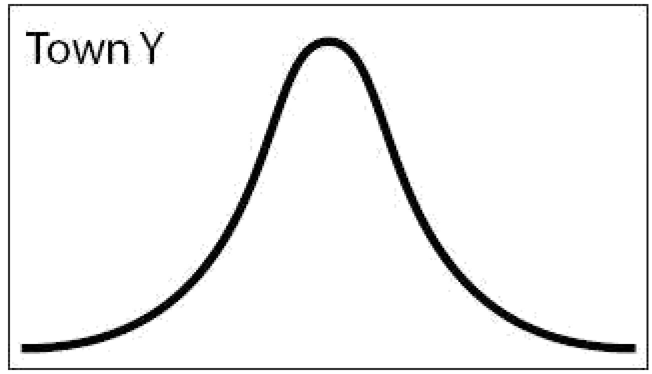

A num ber of scientists' salaries w ere reported;physicists' salaries clustered around a m ean of $100,000,and biologists' clustered around a m ean of $70,000.W hich of the follow ing <u>could</u> be true,according to the graph above?

Indicate <u>all</u> such statem ents.

☐ Som e biologists earn m ore than som e physicists.
☐ B oth biologists' and physicists' salaries are norm ally distributed.
☐ The range of salaries is greater than $150,000

41.



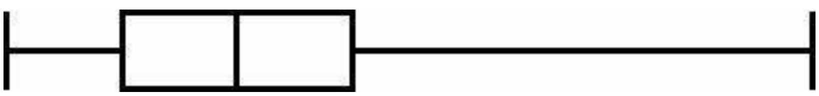




The graph on the left represents the num ber of fam ily m em bers per fam ily in Tow n X ,w hile the graph on the right represents the num ber of fam ily m em bers per fam ily in Tow n Y .The m edian fam ily size for Tow n X is equal to the m edian fam ily size for Tow n Y .The horizontal and vertical dim ensions of the boxes above are identical and correspond to the sam e m easurem ents.W hich of the follow ing m ust be true?

Indicate all such statem ents.

- ☐ The range of fam ily sizes m easured as the num ber of fam ily m em bers is larger in Tow n X than in Tow n Y .
- ☐ Fam ilies in Tow n Y are m ore likely to have sizes w ithin 1 fam ily m em ber of the m ean than are fam ilies in Tow n X .
- ☐ The data for Tow n X has a larger standard deviation than the data for Tow n Y .

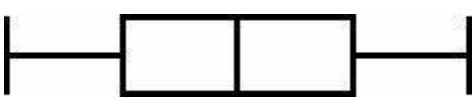
42.



The box-and-w hisker plot show n could be a representation of w hich of the follow ing?

(A ) a data set w ith a range of 100,sym m etrically distributed around its m
edian (B ) a data set w ith a range of 10 and an interquartile range of 6
(C ) a data set in w hich the m edian of the upper half of the data is closer to the low est value in the set than to the highest value
(D ) a set of consecutive integers
(E) a norm al distribution

43.



The box-and-w hisker plot show n could be a representation of w hich of the follow ing sets?

(A ) -2,0,2,4
(B ) 3,3,3,3,3,3
(C ) 1,25,100
(D ) 2,4,8,16,32
(E) 1,13,14,17

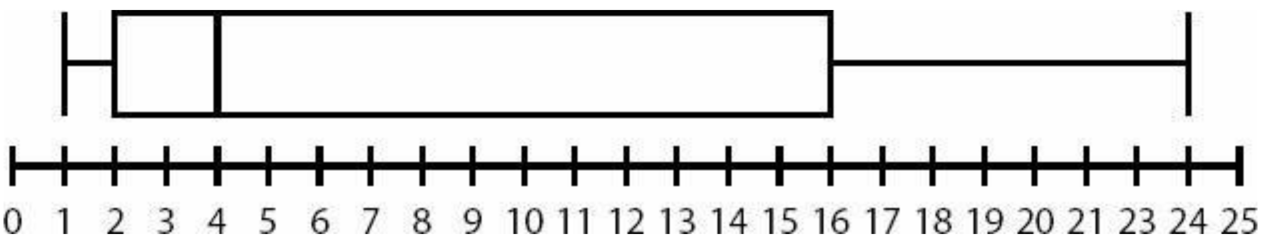
44.

W hich of the follow ing <u>m ust</u> be true about the data described by the box-and-w hisker plot above?

Indicate <u>all</u> such statem ents.

☐ The m edian of the w hole set is closer to the m edian of the low er half of the data than it is to the m edian of the upper half of the data.
☐ The data is norm ally distributed.
☐ The set has a standard deviation greater than zero.

45.




The box-and-w hisker plot above represents a data set w ith:

(A ) a m ean of 4 and a range of 14 (B
) a m ean of 4 and a range of 23 (C ) a
m edian of 4 and a range of 14 (D ) a
m edian of 4 and a range of 23
(E) a m edian of 4 and a range of 24

46.The earthw orm s in Sam ple A have an average length of 2.4 inches,and the earthw orm s in Sam ple B have an average length of 3.8 inches.The average length of the earthw orm s in both sam ples is 3.0 inches.W hich of the follow ing m ust be true?

Indicate <u>all</u> such statem ents.

☐ There are m ore earthw orm s in Sam ple A than in Sam ple B .
☐ The m edian length of the earthw orm s is 3.2 inches.
☐ The range of lengths of the earthw orm s is 1.4.

# Standard D eviation and N orm al D istribution A nsw ers

1.**I and III only.**The w ord *m ean* is a synonym for the average.B ecause an average is calculated by taking the sum of the num bers in the set and dividing by the num ber of num bers in the set,changing *any* one num ber in a set (w ithout adjusting the others) w ill change the sum and,therefore,the average.The m edian is the m iddle num ber in a set,so making the biggest num ber even bigger w on't change that (the m iddle num ber is still 10).Standard deviation is a measure of how *spread out* the num bers in a set are — the m ore spread out the num bers,the larger the standard deviation — so m aking the biggest num ber *really far aw ay* from the others w ould greatly increase the standard deviation.

2.**(D ).**Standard deviation is a m easure of how "spread out" the num bers in a set are — in other w ords,how far are the individual data points from the average of all of the data points? The G R E w ill not ask you to calculate standard deviation — in problem s like this one,you w ill be able to eyeball w hich sets are m ore spread out and w hich are less spread out.

Since Set *Z*'s m em bers are identical,the standard deviation is zero.Zero is the sm allest possible standard deviation for any set,so it m ust be the sm allest here.Y ou can elim inate answ er choices (A ),(B ),and (C ).Set *Y*'s m em bers have a *spread* of 1 betw een each num ber,Set *X* 's m em bers are tw o aw ay from each other,and *W* 's m em bers are six aw ay from each other,so Set *Y* has the next-sm allest standard deviation (note that this is enough to elim inate answ er choice (E) and choose answ er choice (D )).The correct answ er is (D ) *Z*,*Y*,*X* ,*W* .

3.**(B ).**"Set N is a set of *x* distinct positive integers w here $x > 2$" just m eans that the m em bers of the set are all positive integers different from each other,and that there are at least 3 of them .Y ou don't know anything about the standard deviation of the set other than that it is not zero.(B ecause the num bers are different from each other,they are at least a little spread out,w hich m eans the standard deviation m ust be greater than zero.The only w ay to have a standard deviation of zero is to have a set of identical num bers).

In Q uantity B ,m ultiplying each of the distinct integers by -3 w ould definitely spread out the num bers and thus increase the standard deviation.For instance,if the set had been 1,2,3,it w ould now be -3,-6,-9.The negatives are irrelevant — m ultiplying any set of *different* integers by 3 w ill spread them out m ore.

Thus,w hatever the standard deviation is for the set in Q uantity A ,Q uantity B m ust represent a larger standard deviation because the num bers in that set are m ore spread out.

4.**IV only.**In a set of distinct (different) integers,if each num ber is m ultiplied by 1/2,the num bers w ill get closer together (for instance,2,4,6 w ould becom e 1,2,3),so the standard deviation w ould *decrease.*

If the sm allest num ber in a set becam e equal to the m edian,then tw o num bers in the set w ould now be the sam e.The set w ould becom e *less* spread out,not m ore.

If the sm allest num ber w ere increased to becom e larger than the current largest num ber,the standard deviation *could* increase,but w ouldn't have to.For instance,if the set w ere 1,2,3,and the 1 w ere increased to becom e 100,the standard deviation w ould increase.B ut if the set w ere 1,100,101,and the 1 w ere increased to becom e 102,the set w ould get closer together.

Finally,if the largest num ber w ere doubled,the standard deviation w ould have to increase.For instance,if the set w ere 1,2,3 and the largest num ber doubled to 6,then the set w ould becom e m ore spread out.B ecause only the largest num ber is changing,and because the largest num ber becom es even larger w hen doubled,the num bers in the set w ill alw ays becom e m ore spread out,thus increasing the standard deviation.

5.**(D ).**Scoring scales on a test are not necessarily linear,so w hile it m ay look like you can line up the difference in percentiles w ith the difference in score,you cannot m ake any predictions about *other* percentiles.For all you know , 750 could be the 95th percentile score — or 963 could be.A ll you know is that 25% of the scores are $\leq$ 450,w hile 50% of the scores are > 450 and $\leq$ 700,and 25% of the scores are > 700.

6.**I and III only.**For the first statem ent,if the sm aller num ber is positive and 3 less than the larger num ber A N D one term is tw ice the other,the tw o term s have to be 3 and 6.(If you couldn't w ork out the num bers,you could w rite this as tw o equations: $S + 3 = L$ and $L = 2S$.) If you know the num bers,you can calculate the average.

The second statem ent is N O T sufficient.The standard deviation tells you how far all the term s are from the m ean — but know ing *how spread out* the num bers are doesn't tell you w hat they're spread out *from* .For instance,the sets [3, 3,3,6,6,6,6,6,6] and [-3,-3,-3,-6,-6,-6,-6,-6,-6] have the sam e standard deviation and m eet the other constraints of the problem — each set consists of 9 term s,but only tw o different num bers,and 6 of the term s are each tw ice the value of each of the rem aining 3.H ow ever,the tw o sets w ould have different averages (w ithout calculating,you can easily see that one average w ould be positive and one w ould be negative).

The third statem ent w orks — if the biggest term in the set is 6,the sm allest w ould,of course,be 3,allow ing you to calculate the average.

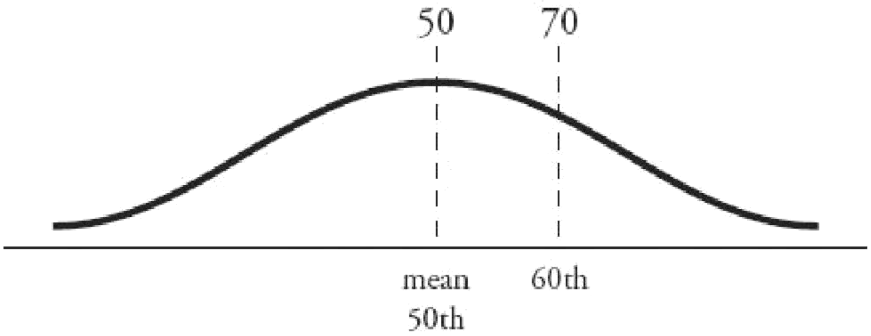
7.**(C ).**A quartile is defined as the m edian of half of a set of data.The first quartile (or Q 1) of a set of data is the m edian of the low er half of the data.For the first half,{2,5,7,11,16,24},the m edian is (7 + 11)/2 = 9 = Q 1.

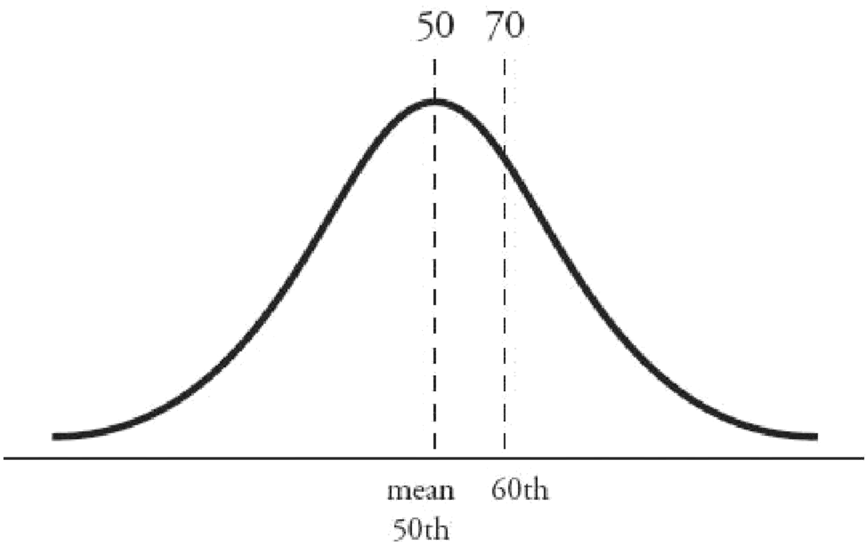
The third quartile (or Q 3) of a set of data is the m edian of the upper half of the data.For the second half,{28,50,52, 101,120,130},the m edian is (52 + 101)/2 = 76.5 = Q 3.

N ow find the average of Q 1 and Q 3 = (9 + 76.5) /2 = 42.75.

8.**(D ).**The test scores are distributed norm ally — that is,in a specific hum p-shaped pattern.That pattern is determ ined by tw o num bers: the m ean (w here the peak of the hum p is) and the standard deviation (a m easure of the w idth of the hum p).To know the exact shape of the distribution,you need to know *both* num bers.

It's given that the 60th percentile is equal to a score of 70.The 60th percentile corresponds to a specific point on the right side of the norm al distribution's hum p— the point w here 60% of the area under the curve is to the left and 40% is to the right.H ow ever,you don't know the w idth of the curve.It could be low and flat (im agine a low er m ean,such as 50,and a high standard deviation,w hich allow s the 60th percentile score of 70 to fall far from the m ean),or it could be high and narrow (w ith a m ean of,say,68 and a tiny standard deviation).So a score of 35 (Q uantity B ) could correspond to *any* percentile below 60th,in fact— either less than or greater than the 30th percentile score,w hich is Q uantity A .




9.**(B ).**N orm al distributions are alw ays centered on and sym m etrical around the m ean,so the chance that the w orm ’s length w ill be w ithin a certain 6-centim eter range (or any specific range) is highest w hen that range is centered on the m ean,w hich in this case is 30 centim eters.

M ore specifically,Q uantity A equals the area betw een -2 standard deviations and the m ean of the distribution.In a norm al distribution,roughly 34 + 34 + 14 + 14 = 96% of the sam ple w ill fall w ithin 2 standard deviations above or below the m ean.If you lim it yourself only to the 2 standard deviations below the m ean,then half of that,or 96% /2 = 48% ,w ill fall in this range.In contrast,Q uantity B equals the area betw een -1 standard deviation and +1 standard deviation.In a norm al distribution,roughly 34 + 34 = 68% of the sam ple w ill fall w ithin 1 standard deviation above or below the m ean.68% is larger than 48% ,so Q uantity B is larger than Q uantity A .

B ut you don’t need these exact figures to answ er this question! Picture any bell curve — the area under the “hum p” (that is,centered around the m iddle) is bigger! Thus,it has m ore m em bers of the set (in this case,w orm s) in it.

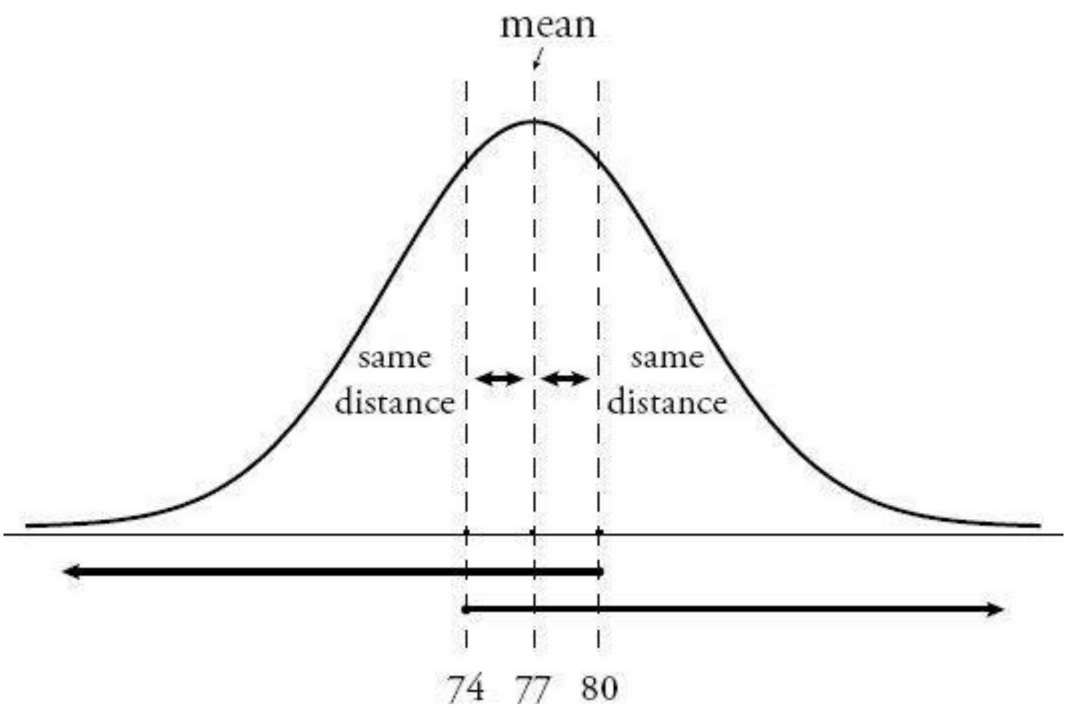
10.**(A ).**The area under a norm al distribution betw een the m ean (the center) and +1 standard deviation represents approxim ately 34% of the total area (this is just a fact to m em orize for the G R E),w hile the area betw een +1 and +2 standard deviations represents approxim ately 14% .The sum of those areas is 34 + 14 = 48% ,so the percent of w ages that fall betw een the m ean ($18) and +2 standard deviations ($18 + $3.50 + $3.50 = $25) is 48% .Thus Q uantity A is larger than Q uantity B .

11.**(B ).**H ow m any standard deviations above $90,000 is $112,000? The difference betw een the tw o num bers is $22,000,w hich is 2 tim es the standard deviation of $11,000.So Q uantity A is really the num ber of hom e values greater than 2 standard deviations above the m ean.

In any norm al distribution,roughly 2% w ill fall m ore than 2 standard deviations above the m ean (this is som ething to m em orize).The value of Q uantity A is roughly (8,000)(0.02) = 160,w hich is definitely sm aller than Q uantity B

(300).

12.**(C ).**The norm al distribution is sym m etrical around the m ean.For any sym m etrical distribution,the m ean equals the m edian (also know n as the 50th percentile).Even though you don’t know the standard deviation,the num ber of students w ho scored *less* than 3 points *above* the m ean (77 + 3 = 80) m ust be the sam e as the num ber of students w ho scored *greater* than 3 points *below* the m ean (77 - 3 = 74).A s long as the boundary scores (80 and 74) are placed sym m etrically around the m ean,you w ill have equal proportions.D raw this if it is at all confusing:




N otice that the tw o conditions overlap and are perfectly sym m etrical.Each num ber consists of a short segm ent betw een it and the 50th percentile m ark,as w ell as half of the students (either above or below the 50th percentile m ark).That is,the “less than 80” category consists of the segm ent betw een 80 and 77,as w ell as all students below the 50th percentile m ark (below 77).The “greater than 74” category consists of the segm ent betw een 74 and 77,as w ell as all students above the 50th percentile m ark (above 77).

13.**(D ).**First,m ake the num bers easier to use.Y ou can m ultiply every num ber by the sam e constant,or m ove the decim al the sam e num ber of places for each num ber.If you m ove the decim al four places,the m ean becom es 1630, the standard deviation becom es 84,and the tw o other num bers becom e 1,546 and 1,756.

N ext,“norm alize” the boundaries you care about.That is,take 1,546 m eters (the low er boundary) and 1,756 m eters (the upper boundary) and convert each of them to a num ber of standard deviations aw ay from the m ean.To do so, subtract the m ean,then divide by the standard deviation.

Low er boundary: 1546 - 1630 = -84

-84 ÷ 84 = -1

So the low er boundary is -1 standard deviation (that is,1 standard deviation less than the m ean).

U pper boundary: 1756 - 1630 = 126

126 ÷ 84 = 1.5

So the upper boundary is 1.5 standard deviations above the m ean.

Y ou need to find the probability that a random variable distributed according to the standard norm al distribution falls betw een -1 and 1.5.

U se the approxim ate areas under the norm al curve.A pproxim ately 34 + 34 = 68% w ill fall w ithin 1 standard deviation above or below the m ean,so 68% accounts for the -1 to 1 portion of the standard deviation.W hat about the portion from 1 to 1.5?

A pproxim ately 14% w ill fall betw een 1 and 2 standard deviations above the m ean.Y ou are not expected to know the exact area betw een 1 and 1.5;how ever,since a norm al distribution has its hum p around 0,*m ore* than half of the area betw een 1 and 2 m ust fall closer to 0 (betw een 1 and 1.5).So the area under the norm al curve betw een 1 and 1.5 m ust be greater than half of the area,or greater than 7% ,but less than the full area,14% .

Put it all together.The area under the norm al curve betw een -1 and 1.5 is approxim ately 68% + (som ething betw een 7% and 14% ).The low er estim ate is 68% + 7% = 75% ,and the upper estim ate is 68% + 14% = 82% .

14.**(B ).**First,com pute the standard deviation from the inform ation given.10.8 pounds is 3 standard deviations m ore than 5.85 pounds,since 10.8 is 2 standard deviations *m ore* than the m ean,and 5.85 is 1 standard deviation *less* than the mean.

10.8 - 5.85 = 4.95 pounds = 3 standard deviations.
4.95 ÷ 3 = 1.65 pounds = 1 standard deviation.

N ow com pute the m ean.Since 5.85 pounds is 1 standard deviation below the m ean,5.85 plus 1 standard deviation equals the m ean:

5.85 + 1.65 = 7.5 pounds = m ean.

Q uantity A : *Tw ice the w eight of a baby 2 standard deviations below the m ean.* N ote that you already know 5.85 is one standard deviation below the m ean;subtract another standard deviation (1.65) in order to reach 2 standard deviations below the m ean,then m ultiply by 2.

= 2 (5.85 - 1.65) = 2 (4.2) = 8.4 pounds.

Q uantity B : *The w eight of a baby 1 standard deviation above the m ean.* A dd 1 standard deviation to the m ean.

= 7.5 + 1.65 = 9.15 pounds.

Q uantity B is larger than Q uantity A .

15.**(E ).**The sm allest num ber in the set is -4,so you can elim inate (D ).The largest num ber in the set is 5,so you can elim inate (B ).

The m edian is -1;now check the m edians of the rem aining answ er choices.The m edian of (A ) is betw een 0 and -2, w hich is -1;(A ) could be the right answ er.The m edian of (C ) is -3,w hich is w rong.The m edian of (E) is betw een -2 and 0,w hich is -1;(E) could also be the right answ er.

Q 1 is -3;check Q 1 for both (A ) and (E).The m edian of the sm aller three num bers (-4,-4,-2) for (A ) is -4,w hich is w rong;you w ant Q 1 to be -3.(E) is the only answ er choice left and you can pick it w ithout checking if you're confident in your previous w ork.H ere's the actual proof: the m edian of the sm aller four num bers (-4,-4,-2,-2) is -3.

16.**(B ).**The interquartile range of a set of data is the distance betw een Q 1 (quartile m arker 1,the m edian of the first half of the set) and Q 3 (quartile m arker 3,the m edian of the second half of the set).

The first ten positive m ultiples of 5 are: 5,10,15,20,25,30,35,40,45,50.Q 1 is the m edian of the first 5 term s,or 15.Q 3 is the m edian of the last 5 term s,or 40.

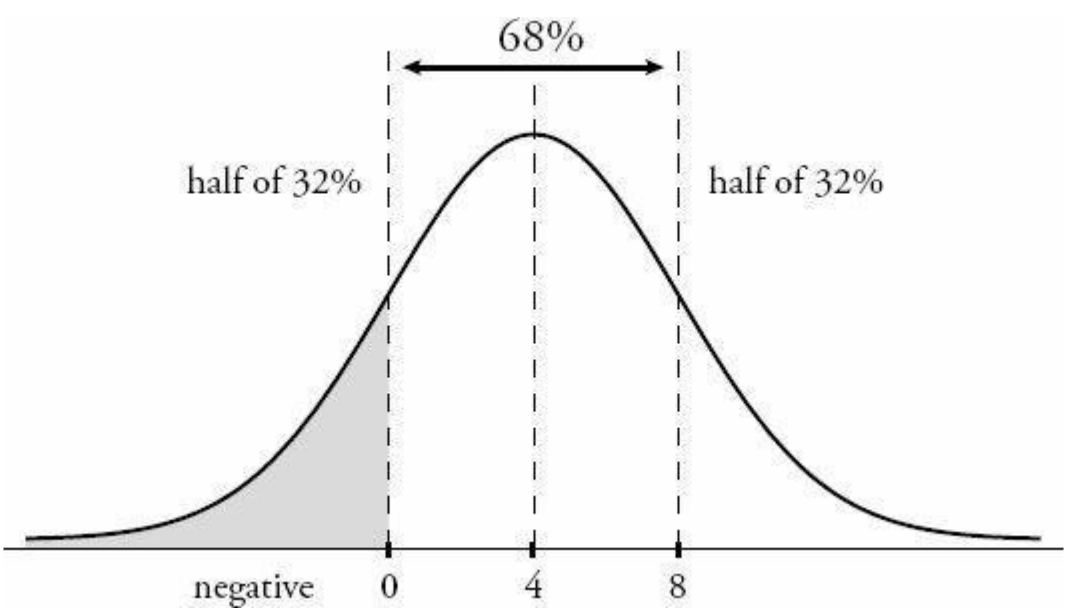
Take the difference betw een Q 3 and Q 1: 40 - 15 = 25.

17.**I only.**The definition of a norm ally distributed set is that about tw o-thirds of the data falls w ithin one standard deviation of the m ean.If only one person scored close to the m ean (and m ost people w ere at the top or bottom of the curve),that set of data is not norm ally distributed,so the first statem ent is true.

The second statem ent is false— the range of the data w ould not necessarily change if the set w ere m ore evenly distributed.For instance,as long as one person still had a zero and one person still had a score of 100,the other scores could fall anyw here w ithout changing the range.

The third statem ent is also false.The m ean of Set *T* m ight or m ight not be equal to the m edian.For instance,the one student w ithin five points of the m ean could have a score actually equal to the m ean;of the rem aining 148 students, half could have scores of 0 and half could have scores of 100.In this case,the m ean w ould equal the m edian. H ow ever,the sam e scenario w ith *unequal* num bers of students scoring 0 and 100 w ould result in the m ean *not* equaling the m edian.

18.**(B ).**In a norm ally distributed set of data,roughly 34 + 34 = 68% of the data fall w ithin one standard deviation of the m ean.This m eans about 100 - 68 = 32% of the data fall outside of one standard deviation,w ith H A LF of that 32% falling one standard deviation to the R IG H T of the m ean,and H A LF of that 32% falling to the left.68%




A negative num ber w ould be m ore than one standard deviation to the left of the m ean (4 - 4 = 0),m eaning that 1/2 of

32%  of the data w ill fall in that range,or 16% .The closest answ er is 1/6.

19.**(C ).**Percentiles define the proportion of a group that scores below a particular benchm ark.Since John scored in the 32nd percentile,by definition,32 percent of the class scored w orse than John.Q uantity A is equal to 32% .

Jane is in the 68th percentile,so 68% of the class scored w orse than she did.Since 100 - 68 = 32,32% of the class scored as high as or higher than Jane.Q uantity B is also equal to 32% .

20.**(D ).**To find the range of one standard deviation,add and subtract the standard deviation from the m ean.For tw o standard deviations,take tw ice the value of the standard deviation (7.1 × 2 = 14.2) and both add and subtract it from the m ean.

4.2 - 14.2 = -10
4.2 + 14.2 = 18.4

21.**41.**The set of data from 25 to 48 has 48 - 25 + 1 = 24 term s,so each octile group w ould be m ade up of 24/8 = 3 term s.It’s easier to find the 6th octile group by w orking backw ards from the 8th one:

8th octile group: 48,47,46   7th octile group: 45,44,43   6th octile group: 42,41,40

The m edian of the 6th octile group is 41.

22.**(B ).**Since the average is 7,you can use the average form ula to find the sum of the scores in the class.

A verage = Sum ÷ # of
term s 7 = Sum ÷ 20
Sum = 140

N ow you already know that at least one student got every possible score.There are eleven possible scores: 0 + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10.This is an evenly spaced set,so you can calculate the sum by m ultiplying the average of the set by the num ber of term s in the set.The average is (10 + 0)/2 = 5 and the num ber of term s is 11,so the sum of the set is 5 × 11 = 55.If you subtract from the sum you found earlier,the rem aining 9 students had to score 140 - 55 = 85 points.

Q uantity B is asking you to find the low est score that tw o students could have received.If 9 students scored a total of 85 points,and any one student could not score m ore than 10 points,then w hat is the low est possible score that any one student could have received? In order to m inim ize that num ber,you need to m axim ize the num bers for the other students.If 8 students scored 10 points each,for a total of 80 points,then the 9th student m ust have scored at least 5. Q uantity B m ust be larger than Q uantity A .N otice that the average score of 7 forces you to m ake a lot of scores 10’s to balance out the very low scores of 0,1,2,etc.that you m ust have in the class (at least one of each).

23.**(B ).**Frequency refers to the num ber of tim es a particular result occurred.In other w ords,if the set represented by this table w ere w ritten out,it w ould look like this: 4,4,4,4,4,4,6,6,6,6,6,8,8,8,8,8.The m ode is the m ost com m only occurring result.The m ode of this data is 4 because 4 appears m ore often than any other num ber.Thus,Q uantity B is larger.

24.**(D ).**Q uintiles (“fifths” of the data) define relative scores,not absolute scores.Im agine tw o possible score distributions:

Exam ple 1: The class’s scores are 1,2,3,4,5 (20% of the class scored each of these).In this case,adding up tw o low est quintile students w ould be 1 + 1 = 2,w hich is less than 5,the score of a top quintile student.

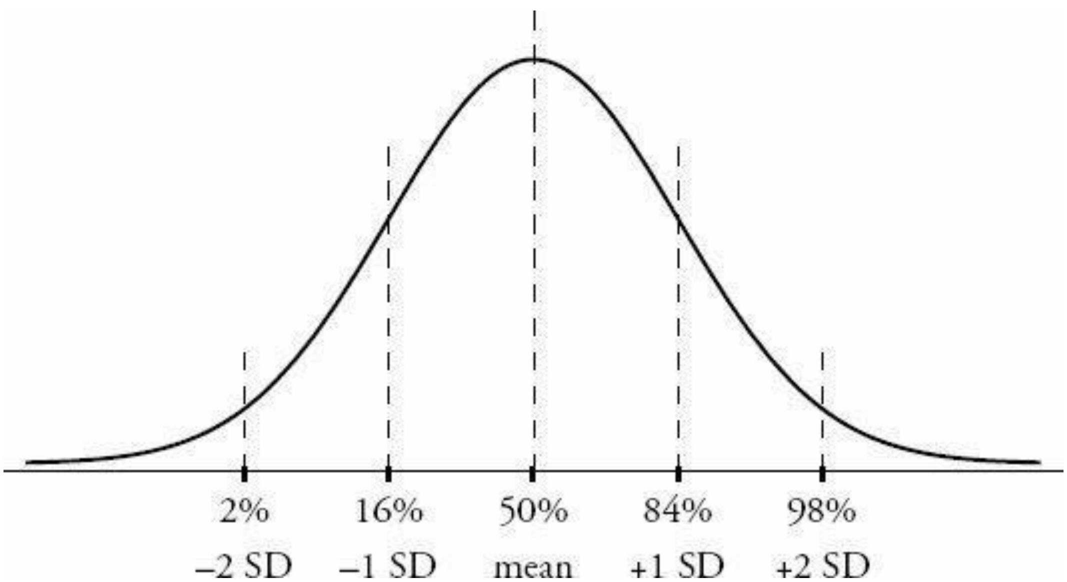
Exam ple 2: The class’s scores are 10,11,12,13,14 (20% of the class scored each of these — still not so sharp!).In this case,adding up tw o low est quintile students w ould be 10 + 10 = 20,w hich is greater than 14,the score of a top quintile student.

Thus,you cannot determ ine w hich quantity is larger.

25.**(E ).**A percentile A LW A Y S represents one *percent* of a set of data.If the question had asked how m any TER M S one percentile represented,that w ould be a different question (w ith a different answ er).

26.**(D ).**Y ou don’t know w hat *x* or *y* are,but since they are both positive integers,they can only be 1 and 18,2 and 9, or 3 and 6 (because they have to m ultiply to 18).So the sm allest num ber in the set is 1 and the largest is 20.Since 20 - 1 = 19,the range is 19.

27.**(D ).**The diagram below show s the standard distribution curve for any norm ally distributed variable.The percent figures correspond roughly to the standard percentiles both 1 and 2 standard deviations (SD ) aw ay from the m ean.




The 2nd percentile is 1720,roughly corresponding to 2 standard deviations below the m ean.Therefore,the m ean - 2 standard deviations = 1720.

Likew ise,the 84th percentile is 1990.84% of a norm ally distributed set of data falls below the m ean + 1 standard deviation,so the m ean + 1 standard deviation = 1990.

C all the m ean *M* and the standard deviation *S*.Y ou can now solve for these variables:

$$M - 2S = 1720$$
$$M + S = 1990$$

Subtract the first equation from the second equation:

$3S = 270$

$S = 90$

Y ou are looking for the 16th percentile,w hich is the m ean - 1 standard deviation or $M - S$.(It’s a fact to m em orize that approxim ately 2% of norm ally distributed data falls below $M - 2S$,and approxim ately 14% of norm ally distributed data falls betw een $M - 2S$ and $M - S$.)

Y ou already know that $M - 2S = 1720$,so add another $S$ to get $M - S$.

$$(M - 2S) + S = 1720 + 90 = 1810$$

N otice that the percentiles are *not* linearly spaced.The norm al distribution is hum p-shaped,so percentiles w ill be bunched up around the hum p and spread out farther aw ay.

28.**(D ).**In m ost data sets,the range is larger than the interquartile range because the interquartile range ignores the sm allest and largest data points.That’s actually the purpose of interquartile range — to get a good picture of w here *m ost* of the data is (think of the “big hum p” on a bell curve).For instance:

Exam ple Set A : 1,2,3,4,5,6,7,100

H ere,the range is 100 - 1 = 99.

The interquartile range is Q 3 - Q 1,or the m edian of the upper half of the data m inus the m edian of the low er half of the data: 6.5 - 2.5 = 4.

In this exam ple,the range is m uch larger.H ow ever,consider this set:

Exam ple Set B : 4,4,4,4,5,5,5,5

In this set,the range is 5 - 4 = 1.The interquartile range is also 5 - 4 = 1.W hile the interquartile range can never be *sm aller* than the range,they can certainly be equal.

29.**67.**Since the standard deviation is 10 and 1 standard deviation above the m ean is 77,sim ply subtract 77 - 10 = 67 to get the m ean.

30.**(C ).**Since one standard deviation below the m ean is 250 and one standard deviation above the m ean is 420,the m ean/m edian m ust be halfw ay in betw een.Since 420 - 250 = 170 and half of 170 is 85,sim ply add 85 to 250 (or subtract it from 420) to get the m ean/m edian of 335.(N ote that in a norm al distribution,the m ean is equal to the m edian,so the tw o term s can be used interchangeably.)

31.**(A ).**If the standard deviation w ere 3,then one standard deviation below the m ean w ould be 9 and one standard deviation above the m ean w ould be 15,so about 2/3 (m ore precisely 68% ) of the data w ould be betw een 9 and 15 (in a norm al distribution,it is alw ays the case that about 2/3 of the data is w ithin one standard deviation of the m ean).

Since the *actual* standard deviation is *less than* 3,about 2/3 of the data is found w ithin an *even sm aller range* than 9 to 15.For instance,the standard deviation could be 1,and then about 2/3 of the data w ould be betw een 11 and 13.O r the standard deviation could be 2.5,and then about 2/3 of all the data w ould be found betw een 9.5 and 14.5.

Since 2/3 of the data is found w ithin an *even sm aller range* than 9 to 15,the range from 9 to 15 contains *m ore than* 2/3 of the data,so it definitely contains m ore than 60% of all the data points.

D on’t be confused by the use of “num ber of data points.” W hile you don’t know the actual total num ber of data points, you can definitively conclude that Q uantity A is equal to a larger percentage of that total than is Q uantity B .

32.**(A ).**Standard deviation is a m easure of the data’s spread from the m ean.W hile the tw o sets have the sam e range (30 - 10 = 20),they do N O T have the sam e spread.The four extra term s in Q uantity B are identical to the m ean, meaning that,on average,the data in Q uantity B is closer to the m ean than the data in Q uantity A .Thus,Q uantity A is more spread out,on average,and has the larger standard deviation.Y ou do not have to com pute the actual standard deviations to find the answ er here.

33.**(D ).**Standard deviation depends on the difference betw een each num ber in a set and the average (arithm etic m ean) of the set.

Q uantity B is know n,since the average of 3 consecutive m ultiples of 3 (such as 3,6,9) is the m iddle num ber (6),and the differences betw een each of the num bers in the set and that average are therefore know n (-3,0,and 3).W hile you don’t need to calculate the actual standard deviation for the G R E,it certainly can be done: standard deviation is the square root of the average of the squares of these differences,or

$$\sqrt{\frac{(-3)^2+(0)^2+(3)^2}{3}}=\sqrt{\frac{18}{3}}=\sqrt{6}$$

,w hich is betw een 2 and 3.

Q uantity A is *not* know n,and in fact it can be greater or less than Q uantity B .A gain,you don’t have to calculate standard deviation;you should just realize that the num bers in Q uantity A could be m ore or less spread out than the num bers in Q uantity B .Therefore,the answ er is (D ).

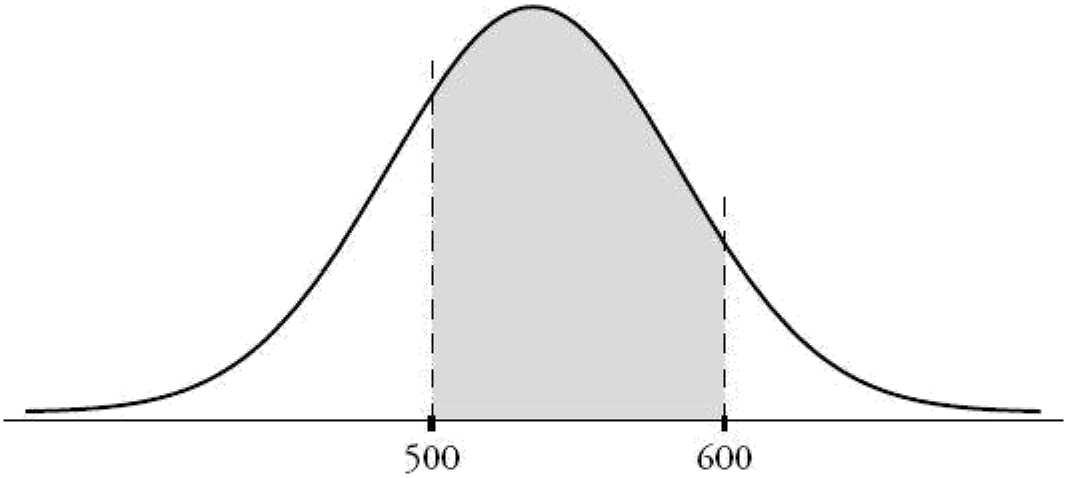
If you’re interested in the calculations: the m axim um value of Q uantity A is 4,w hich you can get from a set of 2 num bers that differ by 8 (say,1 and 9).The average of this set is 5;the difference of each num ber from this average is -4 and 4,and so the standard deviation w orks out to be

$$\sqrt{\frac{(-4)^2+(4)^2}{2}}=\sqrt{\frac{32}{2}}=\sqrt{16}=4$$

.Even w ithout calculating,you should recognized that the set {1,9} is m ore spread out than the set {3,6,9}.

O n the other hand,the standard deviation of a set w ith range of 8 can be as close to zero as you w ant to m ake it.H ow ? Start w ith 1 and 9 again in your set,but now add a w hole bunch of 5’s (right at the average).The differences of those 5’s from the average of the set (w hich is still 5) are all 0.In essence,you’re averaging in a lot of 0’s to the standard deviation,sw am ping the effect of the tw o original deviations from 5 (the original num bers 1 and 9).Y ou can m ake the w hole set be super-close to the m ean,5,reducing the standard deviation to m inuscule levels— w ell below that of the set {3,6,9}.

34.**(D ).**W hile the shaded area m ay appear to be evenly located on either side of the m ean,it isn’t necessarily.For exam ple,the 68% could be m ore lopsided,like so:




This area could still represent 68% of the scores,even if it’s not one standard deviation to either side of the m ean.For you to know that the m edian = 500,the problem w ould need to state explicitly that 500 and 600 each represent one standard deviation from the m ean (or at least that 500 and 600 are equally far from the m ean).

The fact that 68% of the data is located betw een 500 and 600 is definitely intended to trick you into assum ing that 500 and 600 are -1 and +1 standard deviation from the m ean,but you cannot assum e this.W hile it is alw ays true that,in a norm al distribution,about 68% (som e people m em orize the approxim ation as tw o-thirds) of the data is w ithin one standard deviation of the m ean,the reverse is not true: you cannot assum e that any chunk of data that is about 68% of the w hole is therefore w ithin one standard deviation of the m ean.

35.**I and III only**.

I.True.Standard deviation describes how m uch a set of data diverges from the m ean.C urve B is m ore w idely spread than curve A ,and thus *Y* has a larger standard deviation than *X* .

II.N ot True.The probability that *any* norm ally distributed variable falls w ithin 2 standard deviations of its m ean is the sam e,approxim ately 0.14 + 0.34 + 0.34 + 0.14 = 0.96,or 96% .This is a value you should m em orize for the G R E .

III.True.The m ean of a norm al curve is the point along the horizontal axis below the “peak” of the curve.The highest point of curve B is clearly to the right of the highest point of curve A ,so the m ean of Y is larger than the m ean of X . N otice that the m ean has nothing to do w ith the *height* of the norm al curve,w hich only corresponds to how tightly the variable is gathered around the m ean (i.e.,how sm all the standard deviation is).

36.**(C ).**Since the num ber of test results is divisible by 100,the percentiles cleanly divide the total into 100 percentile groups of 300 ÷ 100 = 3 results each.That is,there are 3 results in each percentile.So 2 other results are in the sam e percentile as D om inick’s.Q uantity A is 2.

D om inick’s result is clearly in the 80th percentile,not the 79th or the 81st.So it m ust be possible to distinguish the 80th percentile (that group of 3) from the “next-door” percentiles.Say D om inick got a 58.H ow m any *other* people could have gotten a 58? M aybe no one,m aybe one,m aybe tw o — but if *three* other people got a 58,then you’d have a total of four people w ith the sam e result.In that case,it w ould be im possible to assign D om inick’s result *definitively* to the 80th percentile and not the neighboring percentiles.So the m axim um num ber of other test-takers w ith the sam e result as D om inick is 2.Q uantity B is also 2.

37.**(A ).**400 test scores are distributed am ong 50 possible outcom es (integers betw een 151 and 200,inclusive,w hich num ber 200 - 151 + 1 = 50 integers).There is an average of 400 ÷ 50 = 8 scores per integer outcom e,and there are

400 ÷ 100 = 4 scores in each percentile.So,if all the scores w ere com pletely evenly distributed w ith exactly 8 scores per integer,there w ould be 2 percentile groups per integer outcom e (0th and 1st percentiles at 151,2nd and 3rd percentiles at 152,etc.).In that case,all 50 integers from 151 to 200 w ould correspond to m ore than one percentile group.

H ow can you reduce the num ber of integers corresponding to m ore than one percentile group? B unch up the scores. Im agine that everyone gets a 157.Then that integer is the only one that corresponds to m ore than one percentile group (it corresponds to all 100 groups,in fact).H ow ever,you can’t reduce further this w ay.Y ou’re left w ith 1 integer,so the m inim um num ber of integers corresponding to m ore than one percentile group is 1,w hich is Q uantity A .

A s for Q uantity B ,though,a particular integer m ay have *no* percentile groups corresponding to it.In the previous exam ple,if everyone gets a 157,then no one gets a 158,or a 200 for that m atter.So the m inim um num ber of percentile groups corresponding to a score of 200 (or to any other particular score) is 0,w hich is Q uantity B .

38.**(C ).**A tw o-hum ped shape could com e from tw o overlapping norm al distributions w ith different averages.Since the hum p on the right is sm aller,the distribution w ith a higher average should contain less data.O f the possible answ er choices,only (C ) describes such a scenario.

39.**(B ).**The data appears essentially hum p-shaped,but the left tail is cut off.This m eans that there is no data below a certain cutoff.O nly (B ) corresponds to this kind of situation.

40.**I,II,and III.**

I.C ould B e True.A lthough biologists’ salaries cluster around a low er num ber than physicists’ salaries do,you cannot claim that *every* biologist’s salary is low er than *every* physicist’s salary.Som e biologists’ salaries can be high,and som e physicists’ salaries can be low .

II.C ould B e True.N orm al distributions are consistent w ith the hum p shapes you see in the graph.Y ou cannot *prove* that they’re norm al,but you cannot claim they’re definitely not — they certainly *could* be norm al.

III.C ould B e True.From real-w orld norm al distributions of an unknow n am ount of data,there’s no w ay to tell the m axim um or m inim um values of the data.So the range certainly could be m ore than $150,000.

41.**II and III only.**

I.N ot N ecessarily True.R ange is calculated this w ay: *largest value - sm allest value.*From the graphs as show n (assum ing that they do not continue “off screen” left and right),you w ould conclude that the tw o distributions have the *sam e* range,because the distributions are above zero on both the far left and the far right.(In the real w orld,you m ight assum e that the graphs continue off screen,leading to even less confidence about the range of each distribution.)

II.True.The graph on the right (Tow n Y ) has a sm aller standard deviation (it is less spread out around its m ean).So fam ilies in Tow n Y are m ore likely to be w ithin 1 fam ily m em ber of the m ean than fam ilies in Tow n X are.

III.True.The graph on the left is m ore spread out,so it has a larger standard deviation.

42.**(C ).**The plot is not sym m etrical,so you can elim inate (A ),(D ),and (E),w hich w ould all be sym m etrically distributed around the m edian.Y ou can also elim inate (B ),w hich claim s a range of 10 (the distance from w hisker to w hisker) and an interquartile range of 6 (the w idth of the box).Since the distance betw een the w hiskers is actually m uch m ore than tw ice the w idth of the box,(B ) cannot be the answ er.(C ) fits: the m edian of the upper half of the data (the right edge of the box) is closer to the low est value in the set (the left w hisker) than to the highest value (the right w hisker).

43.**(A ).**The plot is sym m etrical,so you can elim inate any non-sym m etrical data sets (such as (C ),(D ),and (E)).In (B ),all the data points are the sam e,so there w ould be no w idth to the box-and-w hisker plot.(A ) is the only rem aining possibility: the data is evenly spaced,leading to equal w idths for each segm ent of the plot,as show n.

44.**III only.**

I.N ot True.The m edian of the w hole set is the line in the m iddle of the box.A s show n,it is closer to the *right* side of the box (the m edian of the upper half of the data) than to the *left* side of the box (the m edian of the low er half of the data) — the opposite of w hat this statem ent claim s.

II.N ot True.This non-sym m etrical plot could never represent a sym m etrical distribution such as the norm al distribution.In fact,a *true* norm al distribution cannot be represented by a box-and-w hisker plot at all,because such a distribution stretches infinitely to the right and to the left,in theory.

III.True.A ny set represented by a box-and-w hisker plot has a standard deviation greater than zero,because the plot displays som e spread in the data.The only set that has a zero standard deviation is a set containing identical data points w ith zero spread betw een them ,such as {3,3,3,3}.

45.**(D ).**The line inside the box alw ays represents the m edian.If the plot is sym m etrical,the m edian equals the m ean, but because this plot is not sym m etrical,there’s no w ay that the m ean is 4.So you can elim inate (A ) and (B ).

The range is the distance betw een the w hiskers.That distance is 24 - 1 = 23.

46.**I only.**Since the overall average length of all the earthw orm s is closer to the average length of earthw orm s in Sam ple A than to the average for Sam ple B ,there are m ore earthw orm s in Sam ple A .

H ow ever,you cannot know anything about the m edian or the range of the data set w ithout the individual values.For instance,the lengths of all the w orm s in Sam ple A could be exactly 2.4,or they could be spread out quite a bit from 2.4.Sim ilarly,the w orm s in Sam ple B could m easure exactly 3.0,or they could have a variety of different lengths that average to 3.0.Thus,the m edian and range could vary quite a bit.